Tutoriat 10 - Beta random variable, Continuous a priori and a posteriori

Contents

1 Continuous Hypotheses, Discrete Data

We now assume that we have data x that can take one or more discrete values, and a continuous parameter θ that determines the distribution from which the data are drawn.

Notation

- Hypotheses: θ
- Data: x
- Prior: $f(\theta)\,d\theta$
- Likelihood: $p(x \mid \theta)$
- Posterior: $f(\theta \mid x)\,d\theta$

Note. We write $d\theta$ to express the prior and posterior as probabilities rather than densities, even though the prior $f(\theta)$ and posterior $f(\theta \mid x)$ are densities.

Example

Suppose $x \sim \text{Binomial}(5, \theta)$. The domain of θ is $[0, 1]$, and the data x can take 6 possible values: $0, 1, \ldots, 5$.

Since x is discrete and θ is continuous, we use $p(x \mid \theta)$ to describe the likelihood. Suppose the prior is $f(\theta) = 2\theta$. We can construct a table of likelihoods, with each column representing a discrete hypothesis about θ.

hypothesis	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
θ	$(1-\theta)^5$	$5\theta(1-\theta)^4$	$10\theta^2(1-\theta)^3$	$10\theta^3(1-\theta)^2$	$5\theta^4(1-\theta)$

Likelihoods

Question. Suppose we observe data $x = 2$. Use this to compute the posterior for the parameter (hypothesis) θ.

Answer. As before, we choose a column from the likelihood table that will be used in the Bayesian update.

Since we work with probabilities rather than densities, we include $f(\theta)\,d\theta$. On the last line instead of θ^2 it's θ^3. In the final column, instead of $\frac{3!\times 3!}{7!}$ it's $\frac{7!}{3!\times 3!}$

hypothesis	prior	likelihood	Bayes numerator	posterior
θ	$f(\theta)\,d\theta$	$p(x=2 \mid \theta)$	$p(x \mid \theta) f(\theta)\,d\theta$	$f(\theta \mid x)\,d\theta = \frac{p(x \mid \theta) f(\theta)\,d\theta}{p(x)}$
θ	$2\theta\,d\theta$	$\binom{5}{2}\theta^2(1-\theta)^3$	$2\binom{5}{2}\theta^3(1-\theta)^3\,d\theta$	$f(\theta \mid x)\,d\theta = \frac{3!\,3!}{7!}\theta^3(1-\theta)^3\,d\theta$
total	1		$p(x) = \int_0^1 2\binom{5}{2}\theta^2(1-\theta)^3\,d\theta = 2\binom{5}{2}\frac{3!\,3!}{7!}$	1

Figure 1: Bayes table for continuous hypotheses, discrete data

Summary

The prior probabilities are hypotheses, and the likelihoods are given. The posterior probability is obtained by dividing the Bayes numerator by the total probability $p(x)$, which is the integral of the Bayes numerator. Dividing by $p(x)$ normalizes the Bayes numerator.

2 Continuous Hypotheses and Continuous Data

When both the hypotheses are continuous, we change the previous example so that the likelihood function uses $f(x \mid \theta)$ instead of $p(x \mid \theta)$. The general form of the Bayesian update table remains the same.

Notation

- Hypotheses: θ
- Data: x
- Prior: $f(\theta)\,d\theta$
- Likelihood: $f(x \mid \theta)\,dx$
- Posterior: $f(\theta \mid x)\,d\theta$

Simplifying Notation

In previous cases, we included dx so that we worked with probabilities instead of densities. When both the data and the hypotheses are continuous, we need both $d\theta$ and dx. To keep things simple, we omit dx from the tables. This is acceptable because the data x are fixed. We keep $d\theta$ because θ varies.

General Bayesian Update Table

First, we show the general table using simplified notation, followed by the table that includes dx. In the first table, the last column should be interpreted as $f(\theta \mid x)\,d\theta$.

hypothesis	**prior**	**Bayes numerator**	**posterior**
θ	$f(\theta)\,d\theta$	$f(x \mid \theta) f(\theta)\,d\theta$	$f(\theta \mid x) = \dfrac{f(x \mid \theta) f(\theta)\,d\theta}{f(x)}$

$$f(x) = \int f(x \mid \theta) f(\theta)\,d\theta$$

Bayesian Update Table with $d\theta$ and dx

hypothesis	prior	Bayes numerator	posterior
θ	$f(\theta)\,d\theta$	$f(x \mid \theta)f(\theta)\,d\theta\,dx$	$f(\theta \mid x)\,d\theta = \dfrac{f(x \mid \theta)f(\theta)\,d\theta}{f(x)}$

$$f(x)\,dx = \int f(x \mid \theta)f(\theta)\,d\theta\,dx$$

Final Summary

The prior probabilities are hypotheses, and the likelihoods are given data. The posterior probability is obtained by dividing the Bayes numerator by the total probability. The total probability $f(x)$ is the integral of the Bayes numerator, and dividing by $f(x)$ normalizes the Bayes numerator.

3 Normal hypothesis, normal data

In an example of Bayesian inference for continuous data, we assume that both the data and the prior distribution are normal. The following example assumes that the variance of the data is known.

Example 3

Suppose we have a data point $x = 5$ drawn from a normal distribution with unknown mean θ and standard deviation 1:

$$x \sim \mathcal{N}(\theta, 1).$$

Assume further that the prior distribution for θ is

$$\theta \sim \mathcal{N}(2, 1).$$

Let x be a single observation.

(a) Construct the Bayesian table with prior, likelihood, and Bayes numerator.

(b) Show that the posterior distribution for θ is normal.

(c) Find the mean and variance of the posterior distribution.

Solution

We construct the table below, using θ rather than $d\theta$ in the notation. The reason for this is that the total probability calculated by integrating over θ does not depend on θ.

The prior pdf is

$$f(\theta) = \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2}.$$

The likelihood function is

$$f(x = 5 \mid \theta) = \frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2}.$$

Multiplying the prior by the likelihood gives the Bayes numerator:

$$\begin{aligned}
\text{prior} \times \text{likelihood} &= \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2} \\
&= \frac{1}{2\pi} e^{-(\theta^2-4\theta+4+\theta^2-10\theta+25)/2} \\
&= \frac{1}{2\pi} e^{-(2\theta^2-14\theta+29)/2} \\
&= \frac{1}{2\pi} e^{-(\theta^2-7\theta+29/2)}
\end{aligned}$$

Completing the square,

$$\theta^2 - 7\theta + \frac{29}{2} = (\theta - 7/2)^2 + 9/4,$$

so the expression becomes

$$\frac{e^{-9/4}}{2\pi} e^{-(\theta-7/2)^2}.$$

Replacing the complicated constant with a simpler constant c_1, we write

$$c_1 e^{-(\theta-7/2)^2}.$$

Bayesian Table

hypothesis	prior	likelihood	Bayes numerator	posterior $f(\theta \mid x=5)\,d\theta$
θ	$f(\theta)\,d\theta$	$f(x=5 \mid \theta)$	$f(x=5 \mid \theta) f(\theta)\,d\theta$	$\frac{f(x=5 \mid \theta) f(\theta)\,d\theta}{f(x=5)}$
θ	$\frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2}\,d\theta$	$\frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2}$	$c_1 e^{-(\theta-7/2)^2}$	$c_2 e^{-(\theta-7/2)^2}$
total	1		$f(x=5) = \int f(x=5 \mid \theta) f(\theta)\,d\theta$	1

Figure 2: Bayes table for a normal a priori and a posteriori

The posterior pdf is therefore that of a normal distribution. Since the exponential term of a normal distribution is

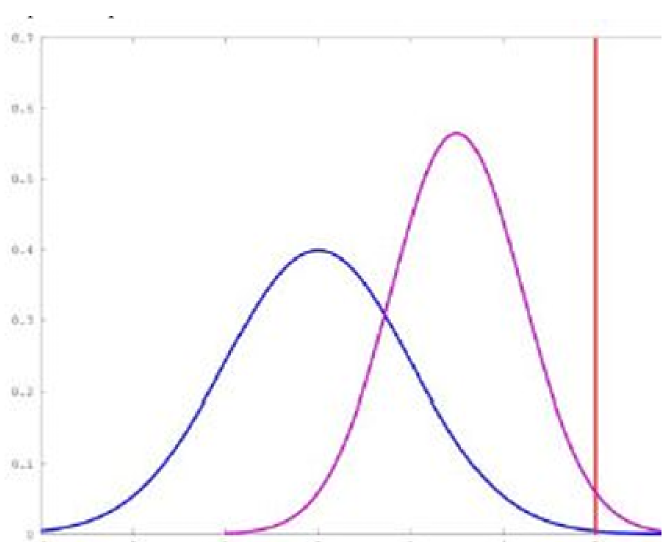
$$e^{-(\theta-\mu)^2/(2\sigma^2)},$$

we identify

$$\mu = \frac{7}{2}, \qquad \sigma^2 = \frac{1}{2}.$$

We do not need to explicitly compute the total probability; it is used only for normalization, and we already know the normalizing constant for the normal distribution.

Graphs of the prior pdf and the posterior pdf illustrate how the observed data "pulls" the prior toward the posterior.



a priori = albastru; a posteriori = mov; data = roşu

Figure 3: Normal distribution prior vs posterior

a priori = blue, a posteriori = purple, data = red

Example 4

Now we repeat the previous example for a general observation x.

Suppose x is drawn from a normal distribution with unknown mean θ and standard deviation 1:

$$x \sim \mathcal{N}(\theta, 1).$$

Solution

The prior pdf and likelihood are

$$f(\theta) = \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2}, \qquad f(x \mid \theta) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}.$$

The Bayes numerator is the product of the prior and the likelihood:

$$\frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2} = \frac{1}{2\pi} e^{-(2\theta^2-2(x+2)\theta+x^2+4)/2}.$$

Completing the square yields

$$c_1 e^{-(\theta-(x+2)/2)^2}.$$

As in the previous example, all constants are absorbed into c_1. In the Bayesian table, we write $f(x)$ in place of $f(x = 5)$.

Thus, the posterior distribution is normal with mean

$$\mu = \frac{x+2}{2}$$

and variance

$$\sigma^2 = \frac{1}{2}.$$

(Compare this result with the case $x = 5$ in the previous example.)

3.1 Predictive Probability Density Functions

Because the data x are continuous, we work with *predictive probability density functions* (pdfs), both prior and posterior.

Prior predictive pdf is the total probability density computed by marginalizing the Bayes numerator:

$$f(x) = \int f(x \mid \theta) f(\theta)\, d\theta,$$

where the integral is taken over the entire domain of θ.

Posterior predictive pdf has the same form as the prior predictive pdf, except that it uses the posterior probabilities for θ:

$$f(x_2 \mid x_1) = \int f(x_2 \mid \theta, x_1) f(\theta \mid x_1)\, d\theta.$$

As usual, we assume that x_1 and x_2 are *conditionally independent*. That is,

$$f(x_2 \mid \theta, x_1) = f(x_2 \mid \theta).$$

In this case, the formula for the posterior predictive pdf becomes slightly simpler:

$$f(x_2 \mid x_1) = \int f(x_2 \mid \theta) f(\theta \mid x_1)\, d\theta.$$

4 The Beta Distribution

The *Beta distribution* $\text{Beta}(a, b)$ is a distribution with **two parameters** and support on the interval $[0, 1]$, with probability density function

$$f(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1}.$$

The following application allows us to explore the shape of the Beta distribution as the parameters vary:

`http://mathlets.org/mathlets/beta-distribution/`

The Beta distribution can be defined for any real numbers $a > 0$ and $b > 0$. Here, we have defined it only for $a, b \in \mathbb{N}^*$, but you can find the full general treatment here:

`http://en.wikipedia.org/wiki/Beta_distribution`

In the context of Bayesian updating, a and b are often called *hyperparameters*, to distinguish them from the unknown parameter θ that represents our hypothesis. In a certain sense, a and b are at a "higher level" than θ, since they parameterize the pdf itself.

4.0.1 A Simple but Important Observation

If a pdf $f(\theta)$ has the form

$$f(\theta) \propto \theta^{a-1}(1-\theta)^{b-1},$$

then $f(\theta)$ is the pdf of a $\text{Beta}(a, b)$ distribution, and the normalization constant must be

$$c = \frac{(a+b-1)!}{(a-1)!(b-1)!}.$$

This result follows because the constant must normalize the pdf so that the total probability is 1. There exists exactly one such constant, and it is given by the formula for the Beta distribution.

A similar observation holds for other distributions such as the normal, exponential, etc.

4.1 Beta Priors and Posteriors for Binomial Random Variables

Example 1

Suppose we have a coin with unknown probability θ of landing heads. We toss the coin 12 times and observe 8 heads and 4 tails. Assuming a uniform prior, show that the posterior pdf is a $\text{Beta}(9, 5)$ distribution.

Solution

Let x_1 denote the data from the 12 tosses. In the table below, we denote by c_2 the constant factor appearing in the posterior column. Our simple observation tells us that this must be the normalization constant of a Beta pdf.

The data consist of 8 heads and 4 tails. Since this is a binomial experiment $\text{Binomial}(12, \theta)$, the likelihood is

$$p(x_1 \mid \theta) = \binom{12}{8}\theta^8(1-\theta)^4.$$

Thus, the Bayesian update table is:

hypothesis	**prior**	**likelihood**	**posterior**
θ	$1 \cdot d\theta$	$\binom{12}{8}\theta^8(1-\theta)^4$	$c_2\,\theta^8(1-\theta)^4\,d\theta$

Our simple observation applies with $a = 9$ and $b = 5$. Therefore, the posterior pdf

$$f(\theta \mid x_1) = c_2\,\theta^8(1-\theta)^4$$

is a $\text{Beta}(9, 5)$ distribution, and the normalization constant must be

$$c_2 = \frac{13!}{8!\,4!}.$$

Note. We explicitly included the binomial coefficient $\binom{12}{8}$ in the likelihood. We could also absorb it into c_2 and not write its explicit value.

Example 2

We now assume that we toss the same coin again, obtaining n heads and m tails. Using the posterior pdf from the previous example as a new prior pdf, show that the new posterior pdf is a Beta$(n+9,\, m+5)$ distribution.

Answer. Everything is in the table. We will denote by x the data from these $n+m$ additional coin tosses. For this data we will no longer write the binomial coefficient explicitly. Instead, we will denote it by c_3. Whenever we need a new constant, we will use c with a new index. Instead of "Bayes posterior" we will read "Bayes numerator," and instead of "numerator" we will read "posterior."

hyp.	prior	likelihood	posterior	numerator
θ	$c_2\,\theta^8(1-\theta)^4$	$c_3\,\theta^n(1-\theta)^m$	$c_2 c_3\,\theta^{n+8}(1-\theta)^{m+4}$	$c_4\,\theta^{n+8}(1-\theta)^{m+4}$
total	1		$T=\int_0^1 c_2 c_3\,\theta^{n+8}(1-\theta)^{m+4}\,d\theta$	1

Again, our simple observation applies, and therefore the posterior pdf

$$f(\theta \mid x_1, x_2) = c_4\,\theta^{n+8}(1-\theta)^{m+4}$$

is that of a Beta$(n+9,\, m+5)$ distribution.

Remark (Flat Beta). The Beta$(1,1)$ distribution coincides with the uniform distribution on $[0,1]$, which we also called a flat prior for θ. This follows by substituting $a=1$ and $b=1$ in the definition of the Beta distribution, which gives $f(\theta)=1$.

4.2 Conjugated a priori

When the the prior and post distributions (let's call it X) are the same for a parameter of another distribution(Y), we say X is a conjugated a priori for Y. As you saw above, the beta distribution is a conjugated a priori for the binomial distribution. .

5 Probability Intervals

Suppose we have a pmf $p(\theta)$ or a pdf $f(\theta)$ describing our belief about the value of an unknown parameter of interest θ.

Definition. A *p-probability interval* for θ is an interval $[a,b]$ such that

$$P(a \le \theta \le b) = p.$$

Remarks.

1. In the discrete case with pmf $p(\theta)$, this means

$$\sum_{a \le \theta_i \le b} p(\theta_i) = p.$$

2. In the continuous case with pdf $f(\theta)$, this means

$$\int_a^b f(\theta)\,d\theta = p.$$

3. We may say *90%-probability interval* for a 0.9-probability interval.

Probability intervals are also called *credible intervals*

Problem

Let

$$X \mid p \sim \text{Binomial}(10, p),$$

where the parameter $p \in (0, 1)$ is unknown. After observing

$$X = 7,$$

compute the probability

$$\mathbb{P}(0.6 \leq p \leq 0.9 \mid X = 7).$$

Assuming a uniform prior on $(0, 1)$, which is the Beta distribution

$$p \sim \text{Beta}(1, 1).$$

Having the density

$$f(p) = 1, \quad 0 < p < 1.$$

Solution

Likelihood

The likelihood of observing $X = 7$ given p is

$$p(X = 7 \mid p) = \binom{10}{7} p^7 (1 - p)^3.$$

—

Bayes' theorem

Bayes' theorem gives the posterior density:

$$f(p \mid X = 7) = \frac{p(X = 7 \mid p)\, f(p)}{p(X = 7)}.$$

Substituting the likelihood and the prior:

$$f(p \mid X = 7) = \frac{\binom{10}{7} p^7 (1 - p)^3 \cdot 1}{p(X = 7)}.$$

—

Compute the normalizing constant

The denominator is the marginal probability of the data:

$$p(X = 7) = \int_0^1 \binom{10}{7} p^7 (1 - p)^3 \, dp.$$

Since the binomial coefficient is constant, we write:

$$p(X = 7) = \binom{10}{7} \int_0^1 p^7 (1 - p)^3 \, dp.$$

Define

$$c_1 = \binom{10}{7}.$$

Then:

$$p(X = 7) = c_1 \int_0^1 p^7 (1 - p)^3 \, dp.$$

The integral is a Beta integral:

$$\int_0^1 p^7(1-p)^3\,dp = B(8,4) = \frac{\Gamma(8)\Gamma(4)}{\Gamma(12)}.$$

Define

$$c_2 = c_1 B(8,4).$$

Thus,

$$p(X=7) = c_2.$$

—

Write the posterior density

Substitute the value of $p(X=7)$ back into Bayes' formula:

$$f(p \mid X=7) = \frac{c_1 p^7(1-p)^3}{c_2}.$$

Define

$$c_3 = \frac{c_1}{c_2}.$$

Then:

$$f(p \mid X=7) = c_3\, p^7(1-p)^3, \quad 0 < p < 1.$$

—

Identify the posterior distribution

Recall the Beta distribution pdf:

$$f(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}.$$

Comparing exponents:

$$a-1 = 7 \Rightarrow a = 8, \quad b-1 = 3 \Rightarrow b = 4.$$

Therefore,

$$p \mid X=7 \sim \text{Beta}(8,4).$$

The normalized posterior density is

$$f(p \mid X=7) = \frac{1}{B(8,4)} p^7(1-p)^3 = \frac{\Gamma(12)}{\Gamma(8)\Gamma(4)} p^7(1-p)^3.$$

—

Compute the desired probability

Finally,

$$\begin{aligned}
\mathbb{P}(0.6 \le p \le 0.9 \mid X=7) &= \int_{0.6}^{0.9} f(p \mid X=7)\,dp \\
&= \int_{0.6}^{0.9} \frac{\Gamma(12)}{\Gamma(8)\Gamma(4)} p^7(1-p)^3\,dp \\
&= I_{0.9}(8,4) - I_{0.6}(8,4) \approx 0.73.
\end{aligned}$$

—

Conclusion

$$\boxed{\mathbb{P}(0.6 \le p \le 0.9 \mid X=7) \approx 0.73}$$

Remark

This probability statement is Bayesian in nature: it represents uncertainty about the parameter p after observing data. Such an interval is called a *credible interval*, not a confidence interval.